Feasibility and Variability of Automated Pupillometry Among Stroke Patients and Healthy Participants: Potential Implications for Clinical Practice.

MARSHALL, Matthew, DEO, Ritesh, CHILDS, Charmaine <http://orcid.org/0000-0002-1558-5633> and ALI, Ali

Published version

MARSHALL, Matthew, DEO, Ritesh, CHILDS, Charmaine and ALI, Ali (2018). Feasibility and Variability of Automated Pupillometry Among Stroke Patients and Healthy Participants: Potential Implications for Clinical Practice. The Journal of neuroscience nursing, p. 1.

Feasibility and Variability of Automated Pupillometry Among Stroke Patients and Healthy Participants – Potential Implications for Clinical Practice

Matthew Marshall[1], Ritesh Deo[1], Charmaine Childs[2], Ali Ali[1]

[1] University of Sheffield, Western Bank, Sheffield S10 2TN

[2] Sheffield Hallam University, Collegiate Campus, Collegiate Crescent, Sheffield, S10 2BP

Key words: Automated pupillometry; Interrater reliability; Neurological observations; Pupillary assessment; Stroke

Word count: 2,143

Number of figures and tables: 2

Corresponding author:

Mr Matthew Marshall, MSc

University of Sheffield

Western Bank

Sheffield S10 2TN

Phone: +447930278505

E-mail: matthew.marshall@kcl.ac.uk

Disclosure statement: The authors report no conflicts of interest

Word Count: 2,143

Abstract

Background: Early neurological deterioration (END) is common after stroke and represents a poor prognostic marker. Manual pupillary assessment to detect END is subjective and has poor interrater reliability. Novel methods of automated pupillometry may be more reliable and accurate. This study aims to evaluate the acceptability and feasibility of automated pupillometry in patients with acute stroke and healthy volunteers, and compare its interrater reliability with the traditional manual method. **Methods:** Automated and manual pupillary assessments were recorded between two independent observers alongside routine neurological observations from twelve acute stroke patients at high risk of END. The proportion of completed measurements, adverse events and qualitative feedback from patients and staff nurses were used to assess acceptability and feasibility of automated pupillometry. Paired automated and manual assessments were supplemented with measures from healthy volunteers to analyse measures of variability and agreement. **Results:** Automated pupillometry was acceptable and safe amongst twelve acute stroke patients but feasibility criteria were not attained. Interrater agreement for automated pupillometry was superior to manual assessment for measurements of pupil size, anisocoria and pupillary light reactivity, for both patients and healthy volunteers. Substantial disparity existed in agreement between automated and manual assessments of these parameters. **Conclusions:** Automated pupillometry represents an alternative to manual pupillary assessment that may have greater interrater agreement and reliability. As an optimised method of neurological assessment it has the potential to improve detection and treatment of conditions leading to END after stroke.

Abstract word count: 237

Introduction

The vast majority of stroke patients in the UK are admitted to a hyper-acute stroke unit (HASU) for urgent recanalising treatments and intensive monitoring for signs of early neurological deterioration (END). END denotes a worsening of an individual's neurological status in their first few days after stroke, affecting approximately 20% of stroke patients and imparting negative prognostic outcomes.[1-3] Close neurological monitoring (Glasgow Coma Scale scores, pupillary responses and vital signs) that occurs on the HASU helps prevent END.[2]

Serial pupillary assessment is a cornerstone of acute neurological monitoring. In the hyper-acute setting, it can sometimes be the only detectable sign of END as sedation, intubation and medications can compromise full neurological assessment.[4] The two main components of the manual pupillary assessment are pupillary light reactivity (PLR) and pupil size. However, these assessments may be confounded by multiple factors. For example, observer bias, subjectivity of terminology, imprecise measuring tools and various external confounders such as ambient light, visual acuity, torch luminosity and iris pigmentation. Moreover, traditional manual methods have poor interrater reliability and thus make clinical interpretation of PLR unreliable.[5-7]

Automation of the pupillary assessment using an automated pupillometer (NPi-200, NeurOptics, Irvine Ca, USA) offers a potential solution to this important clinical problem. The main benefits of automated pupillometry is to produce an objective and standardised assessment of PLR and pupil size. The NPi-200 also provides a Neurological Pupil index (NPi) which produces a scalar value of PLR from 0.0 to 5.0 with a score less than 3.0 being abnormal. This eliminates the subjectivity of traditional PLR assessment and may make trends more clinically meaningful. As automated PLR measures are now being used in a number of settings where neurological assessment is key to detecting END, the aim of this study was to explore the acceptability, feasibility and safety of automated pupillometry after acute stroke. In addition, the interrater reliability of automated pupillometry was compared to the traditional manual method.

Methods

The study was divided into two phases. In the first phase (assessment of feasibility, acceptability and safety), automated and manual pupillary assessments were recorded, alongside routine neurological observations, in stroke patients considered at high risk of END within the first 72 hours of admission to the HASU at Sheffield Teaching Hospitals, UK. Eligible patients included those with any of the following risk factors for END: National Institute of Health Stroke Scale (NIHSS) score > 5, large vessel

occlusion, intracerebral haemorrhage, diabetes, atrial fibrillation and haemorrhagic transformation of infarction or cerebral oedema on initial CT.[2,8,9]

The automated pupillometer would be deemed feasible if at least 80% of readings were completed as intended. The device would be acceptable if all participants and nursing staff rated the assessment positively (1, 2 or 3 out of 5) for comfort and ease of use respectively, this was evaluated using a Likert scale which was as follows: 1 = very comfortable / easy; 2 = comfortable / easy; 3 = borderline; 4 = some discomfort / difficulty; 5 = uncomfortable / very difficult. Safety was defined as the absence of any device-related serious adverse event.

In the second phase (assessment of variability), recordings of paired automated and manual assessments were made to analyse variability and interrater agreement for pupil size, detection of anisocoria and PLR between two independent observers. Agreement between observers for pupil size was defined as a < 1 mm difference between each observer's measurement of the same pupil. Agreement for PLR was defined as both observers recording PLR in the same manual or NPi category as follows: 'non-reactive' (NPi 0.0), 'sluggish' (NPi 0.1 - 2.9), and 'brisk' (NPi 3.0 - 5.0). NPi scores are not directly equivalent to speed of pupillary response. Manual PLR assessment was compared to NPi, in addition to constriction velocity (CV), because NPi scores incorporate multiple additional variables (such as maximum and minimum pupil sizes, percentage change in size, dilatation velocity, and latency of constriction) that may better represent the complexity of the pupillary response.[10] In this study, anisocoria was defined as a ≥ 1 mm difference between pupils, this is consistent with other studies and enabled direct comparison between automated assessment and the whole numbers measured manually.[11] The paired assessments were made within fifteen minutes of each other with observers blind to one another's measurements. Paired measurements were also recorded from healthy volunteers (healthcare staff) to improve the sample size.

Tukey boxplots were utilised to demonstrate the range of measures for PLR and pupil size for manual and automated methods. Percentages of interrater agreement for all three pupillary measures (pupil size, anisocoria and PLR) were calculated between observers. Cohen's kappa coefficient (*k*) values were calculated to ascertain interrater agreement of anisocoria and PLR, and interpreted as follows: poor (< 0.00), slight (0.00 – 0.20), fair (0.21 – 0.40), moderate (0.41 – 0.60), substantial (0.61 – 0.80) and near perfect (0.81 – 1.00). Spearman's correlation coefficient was utilised for pupil size due to the continuous nature of automated values for pupil size.

Results

Twelve stroke patients were recruited during phase 1 of the study (mean age 67.8 years) of whom two thirds were male. 75% of participants had suffered an ischaemic stroke with a median (IQR) NIHSS of 3 (5.75). Diabetes (41.7% of participants) was the most prevalent END risk factor followed by NIHSS > 5 (33.3%), atrial fibrillation (25.0%), cerebral oedema (25.0%), and intracerebral haemorrhage (16.7%). No participants had large vessel occlusion or haemorrhagic transformation on initial CT.

Only 68.4% of intended measurements were completed during the first 72 hours. However, subdividing this result into day and night shifts reveals that 92.7% of daytime measurements (9am to 7pm) were completed as intended compared to 30.8% during the night shift (9pm to 7am). The device was acceptable to both patients (average Likert 1.4 out of 5) and healthcare staff (average Likert 2.4 out of 5), and no device-related serious adverse event occurred.

For analysis of variability and interrater agreement, 132 paired measurements of individual pupils were recorded from a total of 52 participants (42 paired readings from stroke patients and 90 from healthy volunteers). As detection of anisocoria requires comparing both pupils, there were therefore 66 paired measurements for detection of anisocoria.

For automated assessment of pupil size, the interrater agreement was 99.2% whilst the Spearman correlation coefficient was 0.949 (95% CI 0.929-0.969). Using manual methods, interrater agreement of pupil size was 61.4% and the Spearman correlation coefficient was 0.633 (95% CI 0.531-0.735). Figure 1 demonstrates the spread of automated measurements per manually measured pupil size thus giving an indication of agreement between automated and manual methods for assessing pupil size. The vast majority of manual measures were recorded as 3 or 4 mm (84.1%) however the corresponding automated values varied considerably from 1.9 to 6.1 mm.

Using the pupillometer, interrater agreement for detection of anisocoria was substantial (98.5%, k = 0.660, 95% CI 0.039 to 1.00). For manual assessment, interrater agreement was fair (89.4%, k = 0.306, 95% CI -0.078 to 0.690). Interrater agreement among unequal pupils only (n=14) was considerably less at 66.7% and 36.4% for both automated and manual methods respectively. Agreement for detection of anisocoria between automated and manual methods was poor (87.9%, k = -0.027, 95% CI -0.074 to 0.020).

Neither manual nor automated assessments recorded an absent PLR. Interrater agreement of automated assessment of PLR was fair (97.7%, k = 0.389, 95% CI -0.160 to 0.938), while the respective interrater agreement for manual assessment was poor (92.4%, k = -0.039, 95% CI -0.063

to -0.015). Figure 2 demonstrates the lack of agreement between methods for assessment of abnormal PLR. All manually measured sluggish responses were normal when measured using the automated pupillometer (≥ 3.0), and all abnormal NPi scores were marked as brisk manually (k = -0.026, 95% CI -0.042 to -0.010). However, NPi is an indexed function of a number of pupillary measures, and does not compare reaction speeds directly. Mean (SD) CV among pupils manually reported as sluggish was 1.60 m/s (1.08), significantly lower than the CV for those reportedly brisk at 2.51 m/s (0.84, p = 0.001). However, raw agreement between observers for which pupils were reportedly sluggish was 0.0%. Using a definition of abnormal CV as < 0.8 m/s,[10] automated measurements of CV showed that only two of the ten manually reported sluggish pupils had abnormal CV implying only slight agreement between manual measurement of PLR and automated assessment of CV (20.0%, k = 0.006, 95% CI -0.004 to 0.016). In contrast, interrater agreement for automated assessment of CV was perfect (100.0%, k = 1.00, 95% CI 1.00 to 1.00).

Discussion

Automated pupillometry is acceptable and safe, reflecting similar experiences in neurocritical care environments.[12-14] While feasibility criteria were not achieved, this can largely be attributable to poor handover of study information to night staff. Improving the use of automated pupillometry on the HASU can be easily undertaken through improved communication and training of nursing staff.

Automated pupillometry appeared more reliable than manual pupillary assessment in every domain evaluated. Interrater agreement was higher with automated methods for pupil size, anisocoria and PLR. By excluding normal findings to analyse only the occurrences of detected anisocoria and abnormal PLR, automated pupillometry still outperformed manual assessment. However, as the majority of phase 2 participants were healthy volunteers there were relatively few abnormal readings which inhibited more meaningful interpretation. Nonetheless, these findings suggest automated pupillometry is likely to be more reliable than manual methods of pupillary assessment. This is consistent with previous studies evaluating agreement and reliability between automated and manual pupillary assessment in differing neurological populations.[5,12,15]

There was substantial discrepancy between assessment methods as shown in Figures 1 and 2. This disparity implies one method may be less accurate than the other. Meeker et al (2005) found automated pupillometry could detect changes to pupillary dynamics much earlier than manual methods in neurosurgical inpatients.[12] Furthermore, the NPi-200 has been shown to correlate closely to changes in concurrent intracranial pressure (ICP) and may even predict rises in ICP before they are directly measured.[13,14] However, numerous factors other than ICP can influence pupillary dynamics.

The superior accuracy of the NPi-200 is therefore a major assumption that has not been definitively proven.

While manual PLR is not directly comparable to NPi, the fact that none of the ten sluggishly reported pupils had abnormal NPi scores and only two had an abnormal CV suggests that manually reporting sluggish pupils may offer little useful clinical information. This is especially true when considering the use of tracking changes in reactivity from one observer to another as happens during shift handovers of clinicians. Moreover, the NPi-200, as a more precise and comprehensive monitoring tool, is likely to be more sensitive to change and thus tracking changes may help predict END earlier. In contrast, changes in manual pupillary assessment are likely to be large before clinicians can detect changes to act upon. All this evidence suggests manual pupillary assessment could be an unreliable and inaccurate monitoring tool, an imperative notion given its central role within neurological observations.

The findings of this study are directly relevant to clinical practice. Automated pupillometry can be used amongst more alert and conscious stroke patients outside of the intensive care setting. Its use was comfortable, easy to undertake, and devoid of adverse events. Our study suggests that the NPi-200 is a more precise and reliable monitoring tool compared to standard manual assessment. Earlier detection of neurological deterioration may enhance the clinical outcomes for time-dependent interventions (such as decompressive craniotomy) following stroke.

There are disadvantages to automated pupillometry that also need to be considered. The NPi-200 is monocular therefore unable to detect problems in the consensual pupil such as relative afferent pupillary defects. It was also sometimes difficult to perform the assessment in patients unable to follow instructions. This may be a novel limitation of automated pupillometry amongst more alert and conscious patients (as opposed to previous research populations involving more sedated patients). Furthermore, the NPi-200 is considerably more expensive than manual assessment. However, if it was able to accurately differentiate neurological from non-neurological causes of deterioration this may prevent unnecessary and costly neuroimaging. In addition, if automated pupillometry enabled earlier detection of END and referral for life-saving time-dependent treatments, it would be hard not to advocate its value.

There were notable limitations of this study. Primarily, the analysis was not statistically powered to prove superiority of one assessment method over the other. Observer bias may also have influenced manual recordings of pupillary assessment. In addition, the relatively high numbers of healthy participants and consequently limited amount of abnormal measurements mean readers should interpret results with caution.

Conclusion

The NPi-200 appears to be a more reliable and sensitive evaluation of pupillary response compared to manual assessment amongst acute stroke patients and healthy volunteers. The automated pupillometer was being considered for routine clinical use in this HASU. The findings of this study provide underpinning evidence for the integration of automated pupillometry within neurological monitoring after stroke.

References

1. Roquer J, Rodriguez-Campello A, Gomis M et al. Acute stroke unit care and early neurological deterioration in ischaemic stroke. *J Neurol.* 2008;255(7):1012-1017
2. Kwan J, Hand P. Early neurological deterioration in acute stroke: Clinical characteristics and impact on outcome. *Q J Med.* 2006;99(9):625-633
3. Weimar C, Mieck T, Buchthal J et al. 2005. Neurologic worsening during the acute phase of ischemic stroke. *Arch Neurol.* 2005;62(3):393-397
4. Berlin T. Advanced clinical concepts in pupil assessment. 2016
5. Olson D, Stutzman S, Saju C, Wilson M, Zhao W, Aiyagari V. Interrater reliability of pupillary assessments. *Neurocrit Care.* 2016;24(2):251-257
6. Couret D, Boumaza D, Grisotto C et al. Reliability of standard pupillometry practice in neurocritical care: An observational, double-blinded study. *Crit Care.* 2016;20:99
7. Omburo L, Stutzman S, Supnet C, Choate M, Olson D. High Variance in Pupillary Examination Findings Among Postanesthesia Care Unit Nurses. *J Perianesth Nurs.* 2017;32(3):219-224
8. Thanvi B, Treadwell S, Robinson T. Early neurological deterioration in acute ischaemic stroke: Predictors, mechanisms and management. *Postgrad Med J.* 2008;84(994):412-417
9. DeGraba T, Hallenbeck J, Pettigrew K, Dutka A, Kelly B. Progression in acute stroke: Value of the initial NIH stroke scale score on patient stratification in future trials. *Stroke.* 1999;30(6): 1208-1212
10. Shoyombo I, Aiyagari V, Stutzman S et al. Understanding the Relationship Between the Neurologic Pupil Index and Constriction Velocity Values. *Sci Rep.* 2018;8(1):6992
11. Olson D, Fishel M. The use of automated pupillometry in critical care. *Crit Care Nurs Clin North Am.* 2016;28:101-107
12. Meeker M, Du R, Bacchetti P et al. Pupil examination: Validity and clinical utility of an automated pupillometer. *J Neurosci Nurs.* 2005;37(1):34-40
13. McNett M, Moran C, Janki C, Gianakis A. Correlations between hourly pupillometer readings and intracranial pressure values. *J Neurosci Nurs.* 2017;49(4):229-234
14. Chen J, Gombart Z, Rogers S, Gardiner S, Cecil S, Bullock R. Pupillary reactivity as an early indicator of increased intracranial pressure: The introduction of the neurological pupil index. *Surg Neurol Int.* 2011;2:82
15. Zhao W, Stutzman S, Olson D, Saju C, Wilson M, Aiyagari V. Inter-device reliability of the NPi-100 pupillometer. *J Clin Neurosci.* 2016;33:79-82

Figures

Figure 1. Boxplots of automatically measured pupil sizes (using NPi-200) according to each manually measured pupil size.

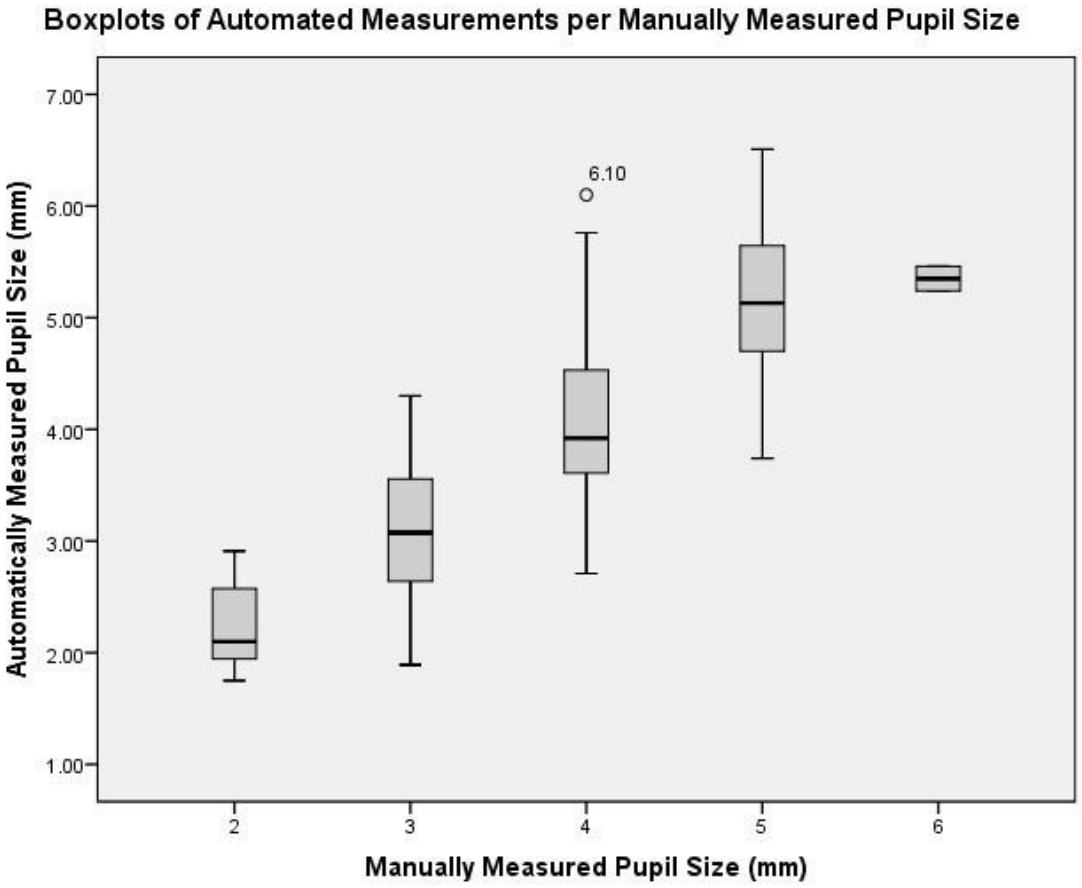

Figure 2. Boxplots of NPi scores according to manually measured sluggish or brisk PLR responses. The dotted line at 3.0 indicates minimum normal NPi score. *NPi = neurological pupil index; PLR = pupillary light reactivity*